⇒MetaMap Home
⇒FAQ Home
⇒Help/Questions

MetaMap's JSON Output

MetaMap JSON output was introduced in MetaMap16V2; previous MetaMap versions do not include this functionality. Basic information about JSON syntax is available here; more detailed information about JSON is available here.

1 Structure of MetaMap's JSON Output

A high-level representation of MetaMap's JSON output for an input file containing three input records or documents is the following:

```
{ "AllDocuments":[

{ "Document": { ...   JSON for Document #1   ... } },
{ "Document": { ...   JSON for Document #2   ... } },
{ "Document": { ...   JSON for Document #3   ... } }

]}
```

2 Generating MetaMap's JSON Output

To generate JSON output, simply call MetaMap with one of these two options:

- `--JSONn`, to generate unformatted JSON output, which contains no formatting whitespace. This form of JSON output is not easily human readable, but creates output files about 40% of the size of MetaMap's formatted JSON, described next.
- `--JSONf N`, to generate formatted JSON output, where `N` is a (preferably small) non-negative integer determining the number of spaces in each incremental indentation. This form of JSON output is human readable, but creates output files over twice as large (assuming `--JSONf 2`) as MetaMap's unformatted JSON.

The calls to `metamap.JSON` should be of the form

```
metamap.JSON --JSONn -E [other MetaMap options]
```

or

```
metamap.JSON --JSONf 2 -E [other MetaMap options]
```

See Appendices A and B at the end of this document for examples of both formatted and unformatted JSON output.

3 Internal Structure of MetaMap JSON Output

MetaMap's JSON output is very similar to MetaMap's XML output, which is described here, and uses many of the same tags. Appendix C below explains the structure of MetaMap's JSON output in great detail.

Appendix A: Formatted MetaMap JSON output

The following is the formatted JSON output generated from the input text `no heart attack (HA)` using the command-line option `--JSONf 2`. The line numbers are referred to in Appendix C; they are included in this document for illustrative purposes only—they are not part of MetaMap's JSON output.

```
1  {"AllDocuments":[
2  {
3     "Document": {
4       "CmdLine": {
5         "Command": "metamap.JSON -L 2015 -Z 2015AB --JSONf 2",
6         "Options": [
7           {
8             "OptName": "lexicon_year",
9             "OptValue": "2015"
10          },
11          {
12            "OptName": "mm_data_year",
13            "OptValue": "2015AB"
14          },
15          {
16            "OptName": "JSONf",
17            "OptValue": "2"
18          },
19          {
20            "OptName": "infile",
21            "OptValue": "user_input"
22          },
23          {
24            "OptName": "outfile",
25            "OptValue": "user_output"
26          }]
27      },
28      "AAs": [
29        {
30          "AAText": "HA",
31          "AAExp": "heart attack",
32          "AATokenNum": "1",
33          "AALen": "2",
34          "AAExpTokenNum": "3",
35          "AAExpLen": "12",
36          "AAStartPos": "17",
37          "AACUIs": ["C0027051"]
38        }],
39      "Negations": [
40        {
41          "NegType": "nega",
42          "NegTrigger": "no",
43          "NegTriggerPIs": [
44            {
45              "StartPos": "0",
46              "Length": "2"
47            }],
48          "NegConcepts": [
49            {
50              "NegConcCUI": "C0027051",
51              "NegConcMatched": "-- Heart Attack"
52            }],
          "NegConcPIs": [
            {
              "StartPos": "3",
              "Length": "12"
            }]
        }],
      "Utterances": [
        {
          "PMID": "00000000",
          "UttSection": "tx",
          "UttNum": "1",
          "UttText": "no heart attack (HA)",
          "UttStartPos": "0",
          "UttLength": "20",
          "Phrases": [
            {
              "PhraseText": "no heart attack",
              "SyntaxUnits": [
                {
                  "SyntaxType": "det",
                  "LexMatch": "no",
                  "InputMatch": "no",
                  "LexCat": "det",
                  "Tokens": ["no"]
                },
                {
                  "SyntaxType": "head",
                  "LexMatch": "heart attack",
                  "InputMatch": "heart attack",
                  "LexCat": "noun",
                  "Tokens": ["heart","attack"]
                }],
              "PhraseStartPos": "0",
              "PhraseLength": "15",
              "Candidates": [],
              "Mappings": [
                {
                  "MappingScore": "-1000",
                  "MappingCandidates": [
                    {
                      "CandidateScore": "-1000",
                      "CandidateCUI": "C0027051",
                      "CandidateMatched": "-- Heart Attack",
                      "CandidatePreferred": "Myocardial Infarction",
                      "MatchedWords": ["heart","attack"],
                      "SemTypes": ["dsyn"],
                      "MatchMaps": [
                        {
                          "TextMatchStart": "1",
                          "TextMatchEnd": "2",
                          "ConcMatchStart": "1",
                          "ConcMatchEnd": "2",
                          "LexVariation": "0"
                        }],
                      "IsHead": "yes",
                      "IsOverMatch": "no",
                      "Sources": ["AIR","AOD","CHV","COSTAR","CSP",
                 "CST","DXP","HPO", "ICD10CM","LCH_NW",
                 "LNC","MEDLINEPLUS","MSH","MTH",
                 "MTHICD9","NCI","NCI_CTCAE","NCI_FDA",
                 "NCI_NICHD", "NDFRT","NLMSubSyn","OMIM",
                 "SNM","SNMI","SNOMEDCT_US","SNOMEDCT_VET"],
           "ConceptPIs": [
            {
             "StartPos": "3",
             "Length": "12"
            }],
           "Status": "0",
           "Negated": "1"
          }]
        }]
      }]
    }]
  }
 }
]}
```

Appendix B: Unformatted MetaMap JSON output

The following is the unformatted JSON output generated from the same input text (`no heart attack (HA)`) using the command-line option `--JSONn`. The actual output file contains only one line, but line breaks have been inserted here for display purposes.

```
{"AllDocuments":[{"Document":{"CmdLine":{"Command":"metamap.JSON -L 2015 -Z 2015AB --JSONn","Opti
ons":[{"OptName":"lexicon_year","OptValue":"2015"},{"OptName":"mm_data_year","OptValue":"2015AB"
},{"OptName":"JSONn"},{"OptName":"infile","OptValue":"user_input"},{"OptName":"outfile","OptValu
e":"user_output"}]},"AAs":[{"AAText":"HA","AAExp":"heart attack","AATokenNum":"1","AALen":"2","A
AExpTokenNum":"3","AAExpLen":"12","AAStartPos":"17","AACUIs":["C0027051"]}],"Negations":[{"NegTy
pe":"nega","NegTrigger":"no","NegTriggerPIs":[{"StartPos":"0","Length":"2"}],"NegConcepts":[{"Ne
gConcCUI":"C0027051","NegConcMatched":"-- Heart Attack"}],"NegConcPIs":[{"StartPos":"3","Length"
:"12"}]}],"Utterances":[{"PMID":"00000000","UttSection":"tx","UttNum":"1","UttText":"no heart at
tack (HA)","UttStartPos":"0","UttLength":"20","Phrases":[{"PhraseText":"no heart attack","Syntax
Units":[{"SyntaxType":"det","LexMatch":"no","InputMatch":"no","LexCat":"det","Tokens":["no"]},{"
SyntaxType":"head","LexMatch":"heart attack","InputMatch":"heart attack","LexCat":"noun","Tokens
":["heart","attack"]}],"PhraseStartPos":"0","PhraseLength":"15","Candidates":[],"Mappings":[{"Ma
ppingScore":"-1000","MappingCandidates":[{"CandidateScore":"-1000","CandidateCUI":"C0027051","Ca
ndidateMatched":"-- Heart Attack","CandidatePreferred":"Myocardial Infarction","MatchedWords":["
heart","attack"],"SemTypes":["dsyn"],"MatchMaps":[{"TextMatchStart":"1","TextMatchEnd":"2","Conc
MatchStart":"1","ConcMatchEnd":"2","LexVariation":"0"}],"IsHead":"yes","IsOverMatch":"no","Sourc
es":["AIR","AOD","CHV","COSTAR","CSP","CST","DXP","HPO","ICD10CM","LCH_NW","LNC","MEDLINEPLUS","
MSH","MTH","MTHICD9","NCI","NCI_CTCAE","NCI_FDA","NCI_NICHD","NDFRT","NLMSubSyn","OMIM","SNM","S
NMI","SNOMEDCT_US","SNOMEDCT_VET"],"ConceptPIs":[{"StartPos":"3","Length":"12"}],"Status":"0","N
egated":"1"}]}]}]}]}]}}]}
```

Appendix C: Structure of MetaMap's JSON output

1 Top-Level

As noted at the beginning of this document, a high-level representation of MetaMap's JSON output for an input file containing three input records or documents is the following:

```
{ "AllDocuments":[

{ "Document": { ...   JSON for Document #1   ... } },
{ "Document": { ...   JSON for Document #2   ... } },
{ "Document": { ...   JSON for Document #3   ... } }

]}
```

We now explain the internal structure of a `Document:{ ... }` JSON pair. In the remainder of this Appendix, line numbers refer to those shown in Appendix A above.

2 The Document Object

The `Document` JSON object contains four pairs:

```
"CmdLine": { ... },
"AAs": [ ... ],
"Negations": [ ... ],
"Utterances": [ ... ]
```

2.1 The `CmdLine` Pair

The `CmdLine` pair spans lines 5–27 in Appendix A and represents the command line used to invoke MetaMap. It contains two pairs:

```
"Command": {  command-line call  },
"Options": [  command-line options  ]
```

The command-line call is the text that was used to invoke MetaMap, e.g.,

```
metamap.JSON -L 2015 -Z 2015AB --JSONf 2
```

The command-line options are rendered as JSON objects such as

```
{
  "OptName": "lexicon_year",            corresponds to "-L"
  "OptValue": "2015"
},
{
  "OptName": "mm_data_year",            corresponds to "-Z"
  "OptValue": "2015AB"
},
{
  "OptName": "JSONf",
  "OptValue": "2"
}
```

The three command-line options above include both a name and a value; options requiring no value, such as `--ignore_word_order` (`-i`) or `--all_derivational_variants` (`-D`) appear as simply

```
{
  "OptName": "all_derivational_variants"
},
{
  "OptName": "ignore_word_order"
},
```

2.2 The `AAs` Pair

The `AAs` pair spans lines 28–38 in Appendix A, and represents the acronyms and abbreviations that MetaMap detected in the text. The only acronym in the text is `HA`, whose expansion is `heart attack`; the data describing this AA are rendered in JSON as

```
{
  "AAText": "HA",                     Text of the acronym
  "AAExp": "heart attack",            Text of the acronym's expansion
  "AATokenNum": "1",                  Number of tokens in the acronym (including whitespace tokens)
  "AALen": "2",                       Number of characters in the acronym
  "AAExpTokenNum": "3",               Number of tokens in the expansion (including whitespace tokens)
  "AAExpLen": "12",                   Number of characters in the expansion
  "AAStartPos": "17",                 Character offset of the beginning of the acronym
  "AACUIs": ["C0027051"]              CUI(s), if any, to which the expansion of the acronym was mapped
}
```

2.3 The `Negations` Pair

The `Negations` pair spans lines 39–58 in Appendix A, and represents the negations that MetaMap detected in the text. MetaMap's negation detection is based on Wendy Chapman's NegEx, and includes some significant extensions, documented here.

The only negated concept in the text is `heart attack`, which is negated by the trigger `no`; the data describing this AA are rendered in JSON as

```
{
  "NegType": "nega",                          Type of negation
  "NegTrigger": "no",                         Negation trigger
  "NegTriggerPIs": [                          Negation trigger character offsets
    {
      "StartPos": "0",                        Negation trigger starting offset
      "Length": "2"                           Negation trigger character length
    }],
  "NegConcepts": [                            Negated concept(s)
    {
      "NegConcCUI": "C0027051",               Negated concept's CUI
      "NegConcMatched": "-- Heart Attack"     Negated concept's string
    }],
  "NegConcPIs": [                             Negated concept character offsets
    {
      "StartPos": "3",                        Negated concept starting offset
      "Length": "12"                          Negated concept character length
    }]
}
```

2.4 The `Utterances` Pair

The `Utterances` pair spans lines 59 through the end of Appendix A, and contains all data generated for the utterances in the input text. Each JSON object representing an utterance has the following form:

```
{
  "PMID": "00000000",                        Document ID, e.g., a PubMed PMID
  "UttSection": "tx",                        Section type
  "UttNum": "1",                             1-based index of the utterance within the text
  "UttText": "no heart attack (HA)",         Text of the utterance
  "UttStartPos": "0",                        Utterance's starting offset
  "UttLength": "20",                         Utterance's character length
  "Phrases": [ ... ]                         See Phrase section, next
}
```

2.4.1 The `Phrases` Pair

The `Phrases` pair spans lines 67 through the end of Appendix A, and contains all data generated for the phrases comprising the enclosing utterance. Each JSON object representing a phrase has the following form:

```
{
  "PhraseText": "no heart attack",      Text of the phrase
  "SyntaxUnits": [ ... ],               Syntactic components of phrase; see below
  "PhraseStartPos": "0",                Phrase's starting offset
  "PhraseLength": "15",                 Phrase's length
  "Candidates": [],                     UMLS candidates detected in phrase
  "Mappings": [ ... ]                   Final mappings generated from phrase
}
```

Note that the value of the `Candidates` pair will be `[]` unless the `--show_candidates` (`-c`) command-line option is used, in which case the candidates will be rendered as shown below in the `MappingCandidates` JSON pair described below in Section 2.4.1.2.

2.4.1.1 The `SyntaxUnits` Pair

The `Phrases` pair spans lines 70–84, and contains data representing the syntactic parse of the phrase. Each JSON object representing a syntactic unit has the following form:

```
{
  "SyntaxType": "head",              Type of syntax unit (head, mod, verb, etc.)
  "LexMatch": "heart attack",        Lexical items in syntax unit
  "InputMatch": "heart attack",      Input words in syntax unit
  "LexCat": "noun",                  Lexical category of words
  "Tokens": ["heart","attack"]       Tokens making up lexical item
}
```

2.4.1.2 The `Mappings` Pair

The `Mappings` pair spans lines 88 through the end of Appendix A, and contains all data representing the UMLS concepts identified by MetaMap that appear in MetaMap's final mappings. This pair contains the data probably of greatest use and interest. Each JSON object representing a mapping has the following form:

```
{
  "MappingScore": "-1000",                              Mapping Score
  "MappingCandidates": [                                List of candidate concepts in mapping
    {
      "CandidateScore": "-1000",                        Concept's Score
      "CandidateCUI": "C0027051",                       Concept's CUI
      "CandidateMatched": "-- Heart Attack",            UMLS string representing concept
      "CandidatePreferred": "Myocardial Infarction",    Concept's preferred name
      "MatchedWords": ["heart","attack"],               Input words matched by concept
      "SemTypes": ["dsyn"],                             UMLS Semantic Type(s) of concept
      "MatchMaps": [                                    Correspondence of words in concept and phrase
        {
          "TextMatchStart": "1",                        Position within phrase of first matching word
          "TextMatchEnd": "2",                          Position within phrase of last matching word
          "ConcMatchStart": "1",                        Position within concept of first matching word
          "ConcMatchEnd": "2",                          Position within concept of last matching word
          "LexVariation": "0"                           Lexical variation between concept and input
        }],
      "IsHead": "yes",                                  Does concept include phrase's head?
      "IsOverMatch": "no",                              Is concept an overmatch?
      "Sources": ["AIR","AOD","CHV","COSTAR","CSP",     UMLS sources of concept
                  "CST","DXP","HPO", "ICD10CM","LCH_NW",
                  "LNC","MEDLINEPLUS","MSH","MTH",
                  "MTHICD9","NCI","NCI_CTCAE","NCI_FDA",
                  "NCI_NICHD", "NDFRT","NLMSubSyn","OMIM",
                  "SNM","SNMI","SNOMEDCT_US","SNOMEDCT_VET"],
      "ConceptPIs": [                                   Character offsets of concept
        {
          "StartPos": "3",                              0-based character offset of concept
          "Length": "12"                                Character length of concept
        }],
      "Status": "0",                           Was concept retained (0), excluded (1), or pruned (2)?
      "Negated": "1"                           Was concept negated (0/1)?
    }]
}
```